

A Kernel Version of Spatial Factor Analysis

Nielsen, Allan A.

Technical University of Denmark, DTU Space – National Space Institute

*Presently located at DTU Informatics – Department of Informatics and Mathematical Modelling,
Richard Petersens Plads, Building 321*

DK-2800 Kgs. Lyngby, Denmark

E-mail: aa@space.dtu.dk, <http://www.imm.dtu.dk/~aa>

Keywords

Orthogonal transformations, dual formulation, Q-mode analysis, kernel substitution, kernel trick.

1 Introduction

Based on work by Pearson [1] in 1901, Hotelling [2] in 1933 introduced principal component analysis (PCA). PCA is often used for general feature generation and linear orthogonalization or compression by dimensionality reduction of correlated multivariate data, see Jolliffe [3] for a comprehensive description of PCA and related techniques. An interesting dilemma in reduction of dimensionality of data is the desire to obtain simplicity for better understanding, visualization and interpretation of the data on the one hand, and the desire to retain sufficient detail for adequate representation on the other hand. Schölkopf et al. [4] introduce kernel PCA. Shawe-Taylor and Cristianini [5] is an excellent reference for kernel methods in general. Bishop [6] and Press et al. [7] describe kernel methods among many other subjects. [10] use kernel PCA to detect change in univariate airborne digital camera images. The kernel version of PCA handles nonlinearities by implicitly transforming data into high (even infinite) dimensional feature space via the kernel function and then performing a linear analysis in that space.

In this paper we shall apply kernel versions of PCA, maximum autocorrelation factor (MAF) [8] analysis to irregularly sampled stream sediment geochemistry data from South Greenland. The 2,097 samples each covering on average 5 km² are analyzed chemically for the content of 41 elements.

2 Principal Component Analysis

Let us consider a data set (for example an image) with n observations (or pixels) and p variables (or spectral bands) organized as a matrix X with n rows and p columns; each column contains measurements over all observations from one variable and each row consists of a vector of measurements x_i^T from p variables for a particular observation

$$(1) \quad X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}.$$

The superscript T denotes the transpose. X is sometimes called the data matrix or the design matrix. Without loss of generality we assume that the variables in the columns of X have mean value zero.

2.1 Primal Formulation

In ordinary (primal also known as R-mode) PCA we analyze the variance-covariance matrix $S = X^T X / (n - 1) = 1 / (n - 1) \sum_{i=1}^n x_i x_i^T$ which is p by p . If $X^T X$ is full rank $r = \min(n, p)$ this will lead to r non-zero eigenvalues λ_i and r orthogonal or mutually conjugate unit length eigenvectors u_i ($u_i^T u_i = 1$) from the eigenvalue problem

$$(2) \quad \frac{1}{n-1} X^T X u_i = \lambda_i u_i.$$

We see that the sign of u_i is arbitrary. To find the principal component scores for an observation x we project x onto the eigenvectors, $x^T u_i$. The variance of these scores is $u_i^T S u_i = \lambda_i u_i^T u_i = \lambda_i$ which is maximized by solving the eigenvalue problem, see appendix with matrix B there equal to the identity matrix.

2.2 Dual Formulation

In the dual formulation (also known as Q-mode analysis) we analyze $XX^T / (n - 1)$ which is n by n and which in image applications can be very large. Multiply both sides of Equation 2 from the left with X

$$(3) \quad \frac{1}{n-1} X X^T (X u_i) = \lambda_i (X u_i)$$

or

$$(4) \quad \frac{1}{n-1} X X^T v_i = \lambda_i v_i$$

with v_i proportional to $X u_i$, $v_i \propto X u_i$, which is normally not normed to unit length if u_i is. Now multiply both sides of Equation 4 from the left with X^T

$$(5) \quad \frac{1}{n-1} X^T X (X^T v_i) = \lambda_i (X^T v_i)$$

to show that $u_i \propto X^T v_i$ is an eigenvector of S with eigenvalue λ_i . We scale these eigenvectors to unit length assuming that v_i are unit vectors ($1 = v_i^T v_i \propto u_i^T X^T X u_i = (n - 1) \lambda_i u_i^T u_i = 1$)

$$(6) \quad u_i = \frac{1}{\sqrt{(n-1)\lambda_i}} X^T v_i.$$

We see that if $X^T X$ is full rank $r = \min(n, p)$, $X^T X / (n - 1)$ and $XX^T / (n - 1)$ have the same r non-zero eigenvalues λ_i and that their eigenvectors are related by $u_i = X^T v_i / \sqrt{(n - 1)\lambda_i}$ and $v_i = X u_i / \sqrt{(n - 1)\lambda_i}$. This result is closely related to the Eckart-Young [11, 12] theorem.

An obvious advantage of the dual formulation is the case where $n < p$. Another advantage even for $n \gg p$ is due to the fact that the elements of the matrix $G = XX^T$, which is known as the Gram¹ matrix, consist of inner products of the multivariate observations in the rows of X , $x_i^T x_j$.

2.3 Kernel Formulation

We now replace x by $\phi(x)$ which maps x nonlinearly into a typically higher dimensional feature space. As an example consider a two-dimensional vector $[z_1 \ z_2]^T$ being mapped into $[z_1 \ z_2 \ z_1^2 \ z_2^2 \ z_1 z_2]^T$. This maps the original two-dimensional vector into a five-dimensional feature space so that for example a linear decision rule becomes general enough to differentiate between all linear and quadratic forms including ellipsoids.

¹named after Danish mathematician Jørgen Pedersen Gram (1850-1916)

The mapping by ϕ takes X into Φ which is an n by q ($q \geq p$) matrix

$$(7) \quad \Phi = \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_n)^T \end{bmatrix}.$$

For the moment we assume that the mappings in the columns of Φ have zero mean. In this higher dimensional feature space $C = \Phi^T \Phi / (n - 1) = 1 / (n - 1) \sum_{i=1}^n \phi(x_i) \phi(x_i)^T$ is the variance-covariance matrix and for PCA we get the primal formulation

$$(8) \quad \frac{1}{n - 1} \Phi^T \Phi u_i = \lambda_i u_i$$

where we have re-used the symbols λ_i and u_i from above.

For the corresponding dual formulation we get

$$(9) \quad \frac{1}{n - 1} \Phi \Phi^T v_i = \lambda_i v_i$$

where we have re-used the symbol v_i from above. As above the non-zero eigenvalues for the primal and the dual formulations are the same and the eigenvectors are related by

$$(10) \quad u_i = \frac{1}{\sqrt{(n - 1)\lambda_i}} \Phi^T v_i$$

and $v_i = \Phi u_i / \sqrt{(n - 1)\lambda_i}$.

Here $\Phi \Phi^T$ plays the same role as the Gram matrix above and has the same size, namely n by n (so introducing the nonlinear mappings in ϕ does not make the eigenvalue problem in Equation 9 bigger).

2.3.1 Kernel Substitution

Applying kernel substitution also known as the kernel trick we replace the inner products $\phi(x_i)^T \phi(x_j)$ in $\Phi \Phi^T$ with a kernel function $\kappa(x_i, x_j) = \kappa_{ij}$ which could have come from some unspecified mapping ϕ . In this way we avoid the explicit mapping ϕ of the original variables. We obtain

$$(11) \quad K v_i = (n - 1) \lambda_i v_i$$

where $K = \Phi \Phi^T$ is an n by n matrix with elements $\kappa(x_i, x_j)$. K is symmetric and must be positive semi-definite, i.e., its eigenvalues are non-negative. Normally the eigenvalue problem is formulated without the factor $n - 1$

$$(12) \quad K v_i = \lambda_i v_i.$$

This gives the same eigenvectors v_i and eigenvalues $n - 1$ times greater. In this case $u_i = \Phi^T v_i / \sqrt{\lambda_i}$ and $v_i = \Phi u_i / \sqrt{\lambda_i}$.

2.3.2 Basic Properties

Several basic properties including the norm in feature space, the distance between observations in feature space, the norm of the mean in feature space, centering to zero mean in feature space, and standardization to unit variance in feature space, may all be expressed in terms of the kernel function without using the mapping by ϕ explicitly. Also the scores calculated by projecting the mapped data onto the primary eigenvectors and the variance of these scores may be expressed by the kernel elements, see [5, 6, 19].

2.3.3 Some Popular Kernels

Popular choices for the kernel function are stationary kernels that depend on the vector difference $x_i - x_j$ only (they are therefore invariant under translation in feature space), $\kappa(x_i, x_j) = \kappa(x_i - x_j)$, and homogeneous kernels also known as radial basis functions (RBFs) that depend on the Euclidean distance between x_i and x_j only, $\kappa(x_i, x_j) = \kappa(\|x_i - x_j\|)$. Some of the most often used RBFs are ($h = \|x_i - x_j\|$)

- multiquadric: $\kappa(h) = (h^2 + h_0^2)^{1/2}$,
- inverse multiquadric: $\kappa(h) = (h^2 + h_0^2)^{-1/2}$,
- Gaussian: $\kappa(h) = \exp(-\frac{1}{2}(h/h_0)^2)$,

where h_0 is a scale parameter to be chosen. Generally, h_0 should be chosen larger than a typical distance between samples and smaller than the size of the study area.

3 Maximum Autocorrelation Factor Analysis

In maximum autocorrelation factor (MAF) analysis first suggested by Switzer and Green [8], we maximize the autocorrelation of linear combinations, $a^T x(r)$, of zero-mean original (spatial) variables, $x(r)$, see also Switzer and Ingebritsen [13], Green et al. [9], Nielsen [14] and Nielsen et al. [15]. $x(r)$ is a multivariate observation at location r and $x(r + \Delta)$ is an observation of the same variables at location $r + \Delta$; Δ is a spatial displacement vector.

3.1 Primal Formulation

The autocovariance R of a linear combination $a^T x(r)$ of multivariate $x(r)$ is

$$(13) \quad R = \text{Cov}\{a^T x(r), a^T x(r + \Delta)\} = a^T \text{Cov}\{x(r), x(r + \Delta)\}a = a^T C_\Delta a$$

where C_Δ is the covariance between $x(r)$ and $x(r + \Delta)$. Introduce the multivariate difference $x_\Delta = x(r) - x(r + \Delta)$ with variance-covariance matrix $S_\Delta = 2S - (C_\Delta + C_\Delta^T)$ where S is the variance-covariance matrix of x defined in Section 2. Since

$$(14) \quad a^T C_\Delta a = (a^T C_\Delta a)^T = a^T C_\Delta^T a = a^T (C_\Delta + C_\Delta^T)a/2$$

we obtain

$$(15) \quad R = a^T (S - S_\Delta/2)a.$$

To get the autocorrelation ρ of the linear combination we divide the covariance by its variance $a^T S a$

$$(16) \quad \rho = 1 - \frac{1}{2} \frac{a^T S_\Delta a}{a^T S a} = 1 - \frac{1}{2} \frac{a^T X_\Delta^T X_\Delta a}{a^T X^T X a}$$

where X is defined in Section 2 and X_Δ is a similar matrix for x_Δ with zero-mean columns. C_Δ above equals $X^T X_\Delta / (n-1)$. To maximize ρ we must minimize the Rayleigh coefficient $a^T X_\Delta^T X_\Delta a / (a^T X^T X a)$ or maximize its inverse. This is done by solving a symmetric generalized eigenvalue problem, see appendix.

3.1.1 Regular and Irregular Spatial Sampling

For regularly sampled spatial data, i.e., ordinary digital image data, often a one-pixel horizontal shift, Δ_h , to obtain $x(r + \Delta_h)$ is used to estimate S_{Δ_h} and a one-pixel vertical shift, Δ_v , to obtain $x(r + \Delta_v)$ is used to estimate S_{Δ_v} . S_{Δ} is then a pool of the two. Alternatively, the two one-pixel shifts may be used to estimate $x_{\Delta} = x(r) - [x(r + \Delta_h) + x(r + \Delta_v)]/2$.

For irregularly sampled spatial data, the difference to the nearest neighbor irrespective of direction may be used. Of course one could include some directional constraint.

For both regularly and irregularly sampled data other possibilities exist [14, 16, 17, 18].

3.2 Dual Formulation and Kernelization

As with the kernel principal component analysis we use the combination of the dual formulation and the kernel trick to obtain an implicit non-linear mapping for the MAF transform. A detailed account of this is given in [19].

4 Case Study: Stream Sediment Geochemistry in South Greenland

In 1979-80 the GGU, the Geological Survey of Greenland (now GEUS, the Geological Survey of Denmark and Greenland), in the so-called Sydurán project collected stream sediment samples from a 10,000 km² area in South Greenland. Sample sites were small active streams with catchment areas of 1-10 km². Samples were sieved at 100 mesh and the undersize was analysed. The present study is based on a dataset with 41 variables and 2,097 samples. Two analytical techniques have been used. The concentrations of Ca, Cu, Fe, Ga, K, Mn, Nb, Ni, Pb, Rb, Sr, Ti, Y, Zn and Zr have been determined by energy-dispersive isotope excited x-ray fluorescence and the concentrations of Au, Ag, As, Ba, Br, Co, Cr, Cs, Hf, Mo, Na, Sb, Sc, Se, Ta, Th, U, W, La, Ce, Nd, Sm, Eu, Tb, Yb and Lu have been determined by instrumental neutron activation analysis. These analyses of the samples are identical to the the ones used in [18] but different from the ones reported in [14, 17].

4.1 Geological Setting

The study area is underlain by a Palaeoproterozoic orogen, the Ketilidian orogen, which consists of three major tectono-stratigraphic units: (1) a northern Border zone of tectonically reworked Archaean gneissic basement overlain by Palaeoproterozoic metasediments and metavolcanics in the north-east, (2) a central zone occupied by a calc-alkaline granitic batholith, and (3) a southern migmatite complex of predominantly Palaeoproterozoic metasediments and metavolcanics intruded by post-tectonic rapakivi type granites, see Figure 1 (left) and [20]. The plate-tectonic setting of the orogen has recently been interpreted in [21]. In Mesoproterozoic times the boundary region between the border and the granite zones was subjected to rifting and intrusions of numerous dykes of basaltic to trachytic compositions as well as of felsic alkaline complexes including carbonatites. The region affected by the alkaline magmas is termed the Gardar province, [22].

5 Results and Discussion

Figure 1 (right) shows the 2,097 sample sites in Southern Greenland in red. The study area is approximately 320 km east-west and 210 km north-south. The Delaunay triangulation is shown in blue.

Figure 2 shows the first three linear PCs (left) and linear MAFs (right) as RGB. Autocorrelations for the first three MAFs are 0.8379, 0.7585 and 0.7279. The linear PCA is correlation matrix based and

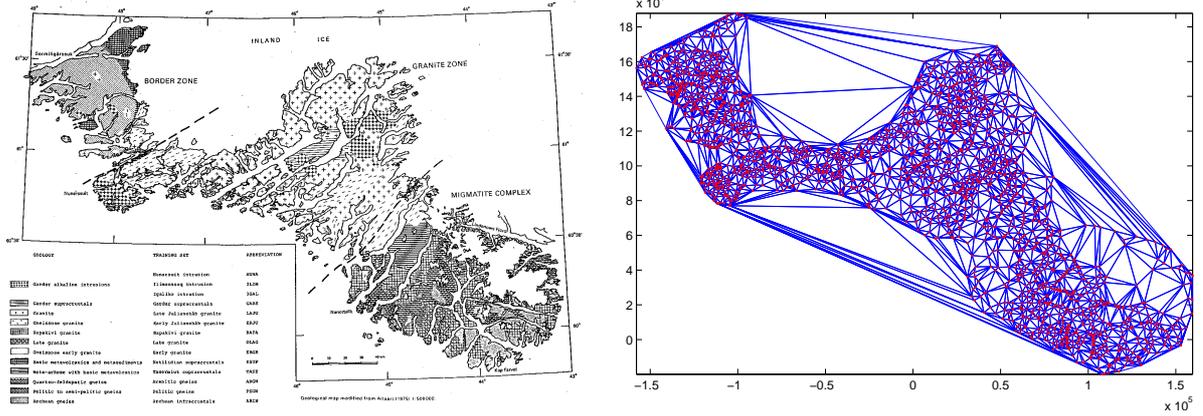


Figure 1: Simplified geological map of South Greenland (left). All 2,097 sample sites and the Delaunay triangulation (right).

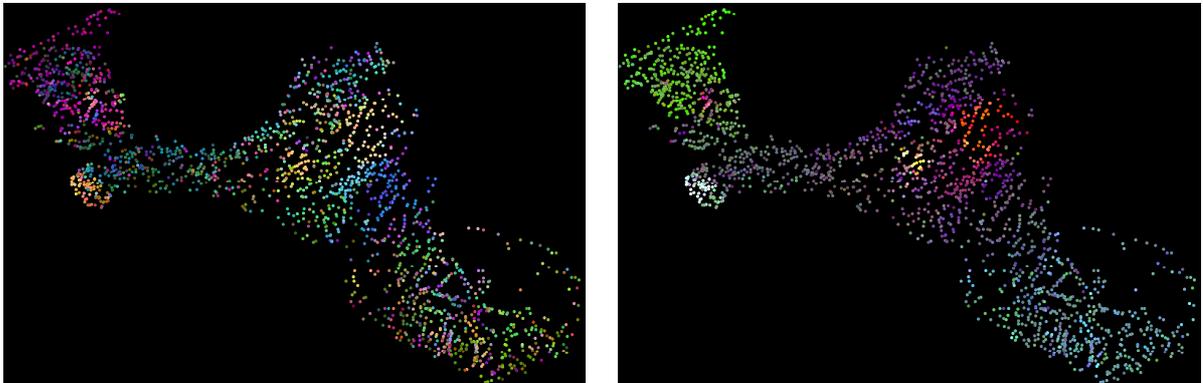


Figure 2: Linear PCs 1, 2 and 3 (left) and linear MAFs 1, 2 and 3 (right) as RGB.

unlike the analyses carried out on the logarithms of the element concentrations in [14, 17, 18], here it is done on the un-transformed data.

Histograms and scatter plots for the 2,097 samples of the 41 elements (not shown) indicate non-Gaussianity and complex correlation patterns. Nonlinear analysis is therefore potentially useful.

For the kernel orthogonalizations a Gaussian kernel $\kappa(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2\sigma^2)$ with σ equal to the mean distance between the observations in feature space is used.

Figure 3 shows 1,850 eigenvalues for kernel PCA (left) and 159 for kernel MAF analysis (right). The 159 largest eigenvalues for kernel MAF analysis correspond to autocorrelations larger than -1 . Autocorrelations for the first three kernel MAFs are 0.9999, 0.9997 and 0.9996, all much higher than achieved by the linear analysis.

Figure 4 shows the first three kernel PCs (left) and kernel MAFs (right) as RGB. Figure 5 shows correlations between original elements and the first three kernel MAFs.

The linear MAF transformation nicely depicts the three major geological units named “Border Zone”, “Granite Zone” and “Migmatite Complex” in the geological map in Figure 1. Also the major alkaline intrusions occurring mainly in the granite zone are clearly depicted. The kernel MAF transformation focusses on the extreme outliers associated with the intrusions and adapts neatly to an even strongly varying background. This behaviour is seen in other application areas also including hyperspectral image data for food quality inspection and change detection studies in both airborne ordinary RGB camera images and hyperspectral scanner Earth observation data.

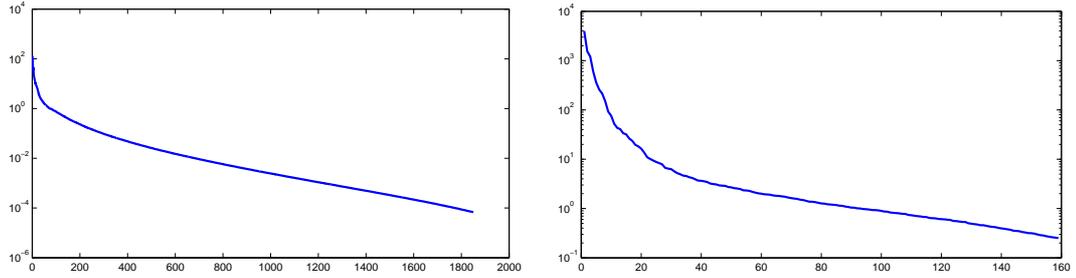


Figure 3: Eigenvalues for kernel PCA (left) and kernel MAF analysis (right).

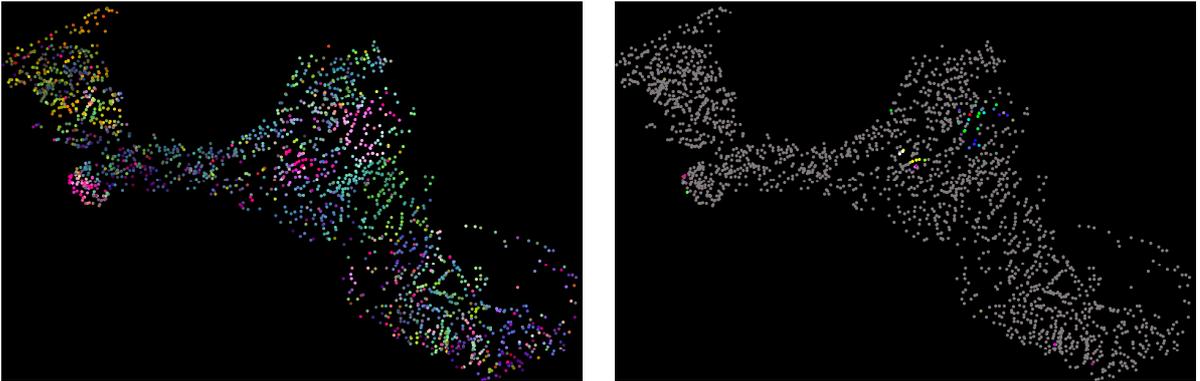


Figure 4: Kernel PCs 1, 2 and 3 (left) and kernel MAFs 1, 2 and 3 (right) as RGB.

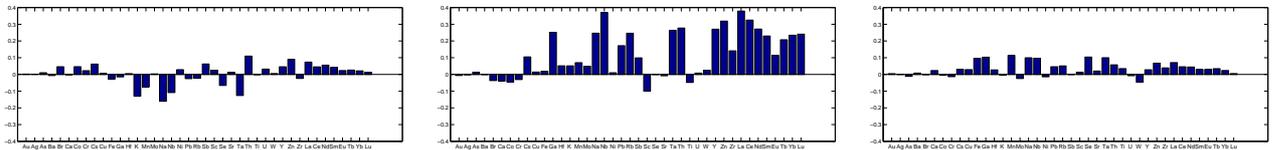


Figure 5: Correlations between original elements and kernel MAFs 1, 2 and 3.

References

- [1] K. Pearson, “On lines and planes of closest fit to systems of points in space,” *Philosophical Magazine*, vol. 6, no. 2, pp. 559–572, 1901.
- [2] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *Journal of Educational Psychology*, vol. 24, pp. 417–441 and 498–520, 1933.
- [3] I. T. Jolliffe, *Principal Component Analysis*, second edition, Springer, 2002.
- [4] B. Schölkopf, A. Smola, and K.-R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [5] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
- [6] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [7] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, third edition, Cambridge University Press, 2007.

- [8] P. Switzer and A. A. Green, "Min/max autocorrelation factors for multivariate spatial imagery," Tech. Rep. 6, Stanford University, 1984.
- [9] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, 1988.
- [10] A. A. Nielsen and M. J. Canty, "Kernel principal component analysis for change detection," Accepted for SPIE Europe Remote Sensing Conference, Cardiff, Great Britain, 2008, Internet <http://www.imm.dtu.dk/pubdb/p.php?5667>.
- [11] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, pp. 211-218, 1936.
- [12] R. M. Johnson, "On a theorem stated by Eckart and Young," *Psychometrika*, vol. 28, no. 3, pp. 259-263, 1963.
- [13] P. Switzer and S. E. Ingebritsen, "Ordering of time-difference data from multispectral imagery," *Remote Sensing of Environment*, vol. 20, pp. 85–94, 1986.
- [14] A. A. Nielsen, "Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data," Ph.D. thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, 1994, Internet <http://www.imm.dtu.dk/pubdb/p.php?296>.
- [15] A. A. Nielsen, K. B. Hilger, O. B. Andersen and P. Knudsen, "A Temporal Extension to Traditional Empirical Orthogonal Function Analysis," In L. Bruzzone and P. Smits (eds.) Proceedings of MultiTemp2001, Trento, Italy, 13-14 September 2001, Analysis of Multi-Temporal Remote Sensing Images, pp. 164–170, 2002. Internet <http://www.imm.dtu.dk/pubdb/p.php?289>.
- [16] A. A. Nielsen, "An Extension to a Filter Implementation of Local Quadratic Surface for Image Noise Estimation," In Proceedings from the 10th International Conference on Image Analysis and Processing, ICIAP'99, pp. 119–124, Venice, Italy, 27-29 September 1999, Internet <http://www.imm.dtu.dk/pubdb/p.php?3937>.
- [17] A. A. Nielsen, K. Conradsen, J. L. Pedersen and A. Steenfelt, "Spatial Factor Analysis of Stream Sediment Geochemistry Data from South Greenland," In V. Pawlowsky-Glahn (ed.) Proceedings of the Third Annual Conference of the International Association for Mathematical Geology, IAMG'97, pp. 955–960, Barcelona, Spain, 22-27 September 1997, Internet <http://www.imm.dtu.dk/pubdb/p.php?5686>.
- [18] A. A. Nielsen, K. Conradsen, J. L. Pedersen and A. Steenfelt, "Maximum Autocorrelation Factorial Kriging", In W. J. Kleingeld and D. G. Krige (eds.) Proceedings of the 6th International Geostatistics Congress, Geostats 2000, pp. 538–547, Cape Town, South Africa, 10-14 April 2000, Internet <http://www.imm.dtu.dk/pubdb/p.php?3639>.
- [19] A. A. Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," Submitted, 2009.
- [20] J. H. Allaart, "Ketilidian mobile belt in South Greenland." In A. Escher and W. S. Watt (Eds.) *Geology of Greenland*. The Geological Survey of Greenland, Copenhagen, pp. 120-151, 1976.
- [21] B. Chadwick and A. A. Garde, "Palaeoproterozoic oblique plate convergence in South Greenland: a reappraisal of the Ketilidian Orogen". In T. S. Brewer (Ed.) *Precambrian Crustal Evolution in the North Atlantic Region*. Geological Society Special Publication No. 112, pp. 179–196, 1996.
- [22] B. G. J. Upton and C. H. Emeleus, "Mid-Proterozoic alkaline magmatism in southern Greenland: Gardar province". In F. G. Fitton and B. G. J. Upton (Eds.) *Alkaline Igneous Rocks*. Geological Society Special Publication No. 30, pp. 449–471, 1987.